Variance Estimates for Price Changes in the Consumer Price Index
January 2004-December 2004

Owen J. Shoemaker

This article presents variance estimates for 1-month, 2-month, 6-month, and 12-month percent changes in the Consumer Price Index for All Urban Consumers (CPI-U). The estimates cover the period January 2004 through December 2004.[1] Each month the CPI collects prices from a sample of approximately 80,000 items in 23,000 outlets around the United States. Variance is a measure of the uncertainty caused by the use of a *sample* of retail prices, instead of the complete universe of retail prices.

The most commonly used measure of sampling variability is the *standard error* of the estimate – the square root of the variance. The standard error of the CPI's change can be used to construct confidence intervals to determine whether the change for a particular CPI series is significantly different from zero. This information should help users determine which index changes are significant.

Presentation of findings

The percent changes in the CPI along with their standard errors were estimated for all 12 months between January 2004 and December 2004. To summarize the results, Tables 1V through 5V show the median values of those percent changes, as well as the median values of their standard errors. Table 1V shows this information for the U.S. City Average, and Tables 2V through 5V show the same information for the Northeast, Midwest, South, and West regions of the country.

For example, from January 2004 through December 2004, the 1-month changes in the U.S. City Average-All Items index had a median value of 0.32 percent. The standard errors of those 12 estimates had a median value of 0.06 percent. Margins of error are usually expressed as a statistic's point estimate plus or minus 2 standard errors, so the margin of error on the CPI's 1-month change is approximately 0.32 percent plus or minus 0.12 percent. That means that in a typical 1-month period the true change in the CPI was probably somewhere between 0.20 percent and 0.44 percent. The tables also show median percent changes and standard errors for 2-month and 6-month intervals and for the full year 2004. Margins of error can be calculated for these intervals in the same way as for 1-month periods.

Analysis of findings

Analysis of the data reveals three important observations. First, the standard errors increase as one moves from the U.S. City Average to individual regions of the country and from *all items* to individual item categories. Second, standard errors differ between item categories. Third, the standard errors decrease on a relative basis (standard error divided by price change), as the price change interval gets longer.

The primary reason standard errors increase as one moves from the U.S. City Average to individual regions of the country is that sample sizes differ. In general, smaller sample sizes lead to larger standard errors. For example, the U.S. City Average-All Items index is computed each month from the prices of approximately 80,000 selected items throughout the United States, and its median standard error for 1-month changes is 0.06 percent. By contrast, the Northeast Region-All Items index is computed from the

[1] In 1998 significant changes were made to the CPI's structure and sample, and a new variance calculation system was implemented. For information on variances from 1978-1986, 1993-1997 and then 1998 and 1999, see the *CPI Detailed Report* for February 1991, May 1994, February 1998, December 1999, and November 2000, respectively.

prices of approximately 17,000 selected items, and its median standard error is 0.18 percent. Regional indexes have larger standard errors because their sample sizes are smaller.

The same effect can be observed as one moves from the all items index to individual item categories. Again, the U.S. City Average-All Items index is computed each month from the prices of approximately 80,000 selected items, and its median standard error is 0.06 percent. By contrast, the U.S. City Average-Recreation index is computed from the prices of approximately 4,000 items, and its median standard error is 0.10 percent, or two-thirds larger. So, again, smaller sample sizes lead to larger standard errors.

The second observation is that standard errors differ between item categories. There are two reasons for this. First, the item categories differ in sample size. For example, the U.S. City Average Food & beverages index is computed from approximately 33,000 prices each month, while the U.S. City Average-Recreation index is computed from approximately 4,000 prices. Therefore, it is not surprising that the recreation index has larger standard errors. Second, there are real differences in item category price behaviors caused by different selling practices, seasonal influences, and consumer demand. This is especially true for the apparel category, in which it is common for the prices of individual items to fluctuate by 50 percent or more each month. As a result, standard errors for apparel indexes are large.

The third observation is that standard errors decrease, on a relative basis (standard error divided by price change), as the price change interval gets longer. For the U.S. City Average - All Items index, the median standard error divided by the median percent change is 0.06/0.32=.188 for 1-month changes, 0.09/0.68=.132 for 2-month changes, 0.12/1.42=.085 for 6-month changes, and 0.12/2.83=.042 for the 12-month change between December 2003 and December 2004. This shows that the relative accuracy of percent changes in the CPI generally improves as the price change interval gets longer. On an absolute basis, the standard errors increase, but at a decreasing rate.

The data presented here indicate that users should exercise caution when using CPI estimates to make inferences about index changes for relatively short time periods, for individual goods and services, or for local areas. The standard errors of those estimates may be on the same order of magnitude as the estimates themselves; and, thus, few inferences about them are reliable.

Sources of error

One way of analyzing the error in a survey estimate is to divide the total error into two sources: *sampling error* and *non-sampling error*. Sampling error is the uncertainty in the CPI caused by the fact that a sample of retail prices is used to compute the CPI, instead of using the complete universe of retail prices. Non-sampling error is the rest of the error. Non-sampling error includes things like incorrect information given by survey respondents, data processing errors, and so forth. Non-sampling error arises, regardless of whether data are collected from a sample of retail prices or from the complete universe.

Another way of analyzing error is to divide it into *variance* and *bias*. The variance of the CPI is a measure of how close different estimates of the CPI would be to each other if it were possible to repeat the survey over and over using different samples. Of course, it is not feasible to repeat the survey over and over, but statistical theory allows the CPI's variance to be estimated anyway. A small variance, for example, indicates that multiple independent samples would produce values that are consistently very close to each other. *Bias* is the difference between the CPI's *expected* value and its *true* value. A statistic may have a small variance but a large bias, or it may have a large variance but a small bias. For an index to be considered accurate, both its variance and bias need to be small.

The Bureau of Labor Statistics is constantly trying to reduce the error in the CPI. Variance and sampling error are reduced by using a sample of retail prices that is as large as possible, given resource constraints. The Bureau has developed a model that optimizes the allocation of resources by indicating the number of prices that should be observed in each geographic area and each item category, in order to minimize the variance of the U.S. City Average-All Items index. The Bureau reduces non-sampling error through a series of computerized and professional data reviews, as well as through continuous survey process improvements and theoretical research.

Replication and variance estimation

An important advantage of using sampling is that the CPI's variance can be estimated directly from the sample data. Starting in 1978, the CPI's sample design has accommodated variance estimation by using two or more independent samples of items and outlets in each geographic area. This allows two or more statistically independent estimates of the index to be made. The independent samples are called *replicates*, and the set of all observed prices is called the *full sample*.

The CPI collects data in 38 geographic areas across the United States. The 38 areas consist of 31 *self-representing* areas and 7 *non-self-representing* areas. Self-representing areas are large metropolitan areas, such as the Boston metropolitan area, the St. Louis metropolitan area, and the San Francisco metropolitan area. Non-self-representing areas are collections of smaller metropolitan areas. For example, one non-self-representing area is a collection of 32 small metropolitan areas in the Northeast region (Buffalo, Hartford, Providence, Bangor, and others), of which 8 were randomly selected to represent the entire set. Within each of the 38 areas, price data are collected for 211 item categories called *item strata*. Together the 211 item strata cover all consumer purchases. Examples of item strata are bananas, women's dresses, and electricity.

Multiplying the number of areas by the number of item strata gives 8,018 (= 38 × 211) different area/item combinations for which price indexes need to be calculated. Separate price indexes are calculated for each one of these 8,018 area/item combinations. After all 8,018 of these *basic-level* indexes are calculated, they are aggregated to form *higher-level* indexes, using expenditure estimates from the Consumer Expenditure Survey as their weights. Examples of higher-level geographic areas are the four regions of the country (Northeast, Midwest, South, and West); and examples of higher-level item categories are the eight major groups (food & beverages, housing, apparel, transportation, medical care, education and communication, recreation, and other goods and services). The highest level of geographic aggregation is the *U.S. City Average*, and the highest level of item aggregation is *All Items*.

Variances are computed with a Stratified Random Groups method, in which variances are computed separately for certain subsets of areas and items and are then combined to produce the variance of the entire area and item combination. Subsets of items are formed by the intersection of the item category with each of the eight major groups.

Let $CPI(A,I,f,t)$ denote the index value for area = A, item category = I, in month = t, where f indicates that it is the full-sample value, and let $CPI(A,I,f,t-k)$ denote the value of the same index in month = $t-k$. In general, the upper-case letter A denotes a *set* of areas, such as the Northeast or Midwest region of the country; and the upper-case letter I denotes a higher-level item category, such as *all items* or a*ll items less food and energy*. Also let $CPI(A,I,r,t)$ and $CPI(A,I,r,t-k)$ be the corresponding index values for replicate = r. Most areas have two replicates, but some have more. Then, the full-sample k-month percent change between months $t-k$ and t is computed by dividing $CPI(A,I,f,t)$ by $CPI(A,I,f,t-k)$, subtracting 1, and multiplying by 100:

$$PC(A,I,f,t,t-k) = \left(\frac{CPI(A,I,f,t)}{CPI(A,I,f,t-k)} - 1 \right) \times 100$$

Every index has a weight *W(A,I,f)* or *W(A,I,r)* associated with it, which is used to combine the index with other indexes to produce indexes for larger geographic areas and larger item categories. For example, the weights are used to combine all 8,018 "basic-level" indexes into higher-level indexes such as the U.S. City Average-All Items index. The product of an index and its weight is called a *cost weight*, $CW(A,I,r,t) = CPI(A,I,r,t) \times W(A,I,r)$, and is an estimate of the total cost in area = A for consumption of item category = I in month = t.

For the Stratified Random Groups method used here, replicate percent changes are defined as follows: full sample cost weights are used for every geographic area within area = A except for one of the areas. In the omitted area, the full sample cost weight is replaced by a replicate cost weight. Let the lower case letter a denote one of the 38 basic-level areas included in area = A, and let the lower case letter i denote the intersection of item category = I with one of the 8 major groups. Then, the replicate percent change, for area = a, item subset = i, replicate = r, between months $t-k$ and t, is computed as:

$$PC_S(a,i,r,t,t-k) = \left(\frac{CW(A,I,f,t) - CW(a,i,f,t) + CW(a,i,r,t)}{CW(A,I,f,t-k) - CW(a,i,f,t-k) + CW(a,i,r,t-k)} - 1 \right) \times 100$$

for self-representing areas. For non-self-representing areas, the replicate percent change, for area = a, item category = I, replicate = r, between months $t-k$ and t, is computed as:

$$PC_N(a,I,r,t,t-k) = \left(\frac{CW(A,I,f,t) - CW(a,I,f,t) + CW(a,I,r,t)}{CW(A,I,f,t-k) - CW(a,I,f,t-k) + CW(a,I,r,t-k)} - 1 \right) \times 100$$

where:

$$CW(A,I,f,t) = \sum_{a \subset A} \sum_{i \subset I} CW(a,i,f,t)$$
$$CW(A,I,f,t) = \sum_{a \subset A} CW(a,I,f,t)$$
$$CW(a,I,f,t) = \sum_{i \subset I} CW(a,i,f,t)$$

and likewise for replicates. The symbol "$a \subset A$" means that the sum is over all basic-level areas within area = A, and the symbol $i \subset I$ means that the sum is over all item categories that are intersections of item category = I with a major group.

Then, the variance is computed with the following Stratified Random Groups variance estimation formula:

$$\begin{aligned} V[PC(A,I,f,t,t-k)] &= \sum_{i \subset I} \sum_{a \subset A \cap S} \frac{1}{R_a(R_a-1)} \sum_{r=1}^{R_a} \left(PC_S(a,i,r,t,t-k) - PC(A,I,t,t-k) \right)^2 \\ &+ \sum_{a \subset A \cap N} \frac{1}{R_a(R_a-1)} \sum_{r=1}^{R_a} \left(PC_N(a,I,r,t,t-k) - PC(A,I,t,t-k) \right)^2 \end{aligned}$$

where S and N are the sets of all self-representing and non-self-representing areas in the CPI's geographic sample, respectively; and $A \cap S$ and $A \cap N$ are the sets of all self-representing and non-self-representing areas within area = A. The number R_a is the number of replicates in area = a.

Finally, the standard error of the percent change is computed by taking the square root of its variance:

$$SE[PC(A, I, f, t, t-k)] = \sqrt{V[PC(A, I, f, t, t-k)]}\ .$$

For more information, write to the CPI Information Office, BLS - Room 3130, 2 Massachusetts Avenue, N.E., Washington, DC 20212, or call Owen Shoemaker at 202-691-6918.